CHAPTER 12

A Model of Heuristic Judgment

Daniel Kahneman
Shane Frederick

The program of research now known as the heuristics and biases approach began with a study of the statistical intuitions of experts, who were found to be excessively confident in the replicability of results from small samples (Tversky & Kahneman, 1971). The persistence of such systematic errors in the intuitions of experts implied that their intuitive judgments may be governed by fundamentally different processes than the slower, more deliberate computations they had been trained to execute.

From its earliest days, the heuristics and biases program was guided by the idea that intuitive judgments occupy a position – perhaps corresponding to evolutionary history – between the automatic parallel operations of perception and the controlled serial operations of reasoning. Intuitive judgments were viewed as an extension of perception to judgment objects that are not currently present, including mental representations that are evoked by language. The mental representations on which intuitive judgments operate are similar to percepts. Indeed, the distinction between perception and judgment is often blurry: The *perception* of a stranger as menacing entails a *prediction* of future harm.

The ancient idea that cognitive processes can be partitioned into two main families – traditionally called intuition and reason – is now widely embraced under the general label of dual-process theories (Chaiken & Trope, 1999; Evans and Over, 1996; Hammond, 1996; Sloman, 1996, 2002; see Evans, Chap. 8). Dual-process models come in many flavors, but all distinguish cognitive operations that are quick and associative from others that are slow and governed by rules (Gilbert, 1999).

To represent intuitive and deliberate reasoning, we borrow the terms "system 1" and "system 2" from Stanovich and West (2002). Although suggesting two autonomous homunculi, such a meaning is not intended. We use the term "system" only as a label for collections of cognitive processes that can be distinguished by their speed, their controllability, and the contents on which they operate. In the particular dual-process model we assume, system 1 quickly proposes intuitive answers to judgment problems as they arise, and system 2 monitors the quality of these proposals, which it may endorse, correct, or override. The judgments that are eventually expressed are called intuitive if they retain the hypothesized initial proposal with little modification.

The effect of concurrent cognitive tasks provides the most useful indication of whether a given mental process belongs to system 1 or system 2. Because the overall capacity for mental effort is limited, effortful processes tend to disrupt each other, whereas effortless processes neither cause nor suffer much interference when combined with other tasks (Kahneman, 1973; Pashler, 1998). It is by this criterion that we assign the monitoring function to system 2: People who are occupied by a demanding mental activity (e.g., attempting to hold in mind several digits) are much more likely to respond to another task by blurting out whatever comes to mind (Gilbert, 1989). By the same criterion, the acquisition of highly skilled performances – whether perceptual or motor – involves the transformation of an activity from effortful (system 2) to effortless (system 1). The proverbial chess master who strolls past a game and quips, "White mates in three" is performing intuitively (Simon & Chase, 1973).

Our views about the two systems are similar to the "correction model" proposed by Gilbert (1989, 1991) and to other dual-process models (Epstein, 1994; Hammond, 1996; Sloman, 1996; see also Shweder, 1977). We assume system 1 and system 2 can be active concurrently, that automatic and controlled cognitive operations compete for the control of overt responses, and that deliberate judgments are likely to remain anchored on initial impressions. We also assume that the contribution of the two systems in determining stated judgments depends on both task features and individual characteristics, including the time available for deliberation (Finucane et al., 2000), mood (Bless et al., 1996; Isen, Nygren, & Ashby, 1988), intelligence (Stanovich & West, 2002), cognitive impulsiveness (Frederick, 2004), and exposure to statistical thinking (Agnoli, 1991; Agnoli & Krantz, 1989; Nisbett et al., 1983).

In the context of a dual-system view, errors of intuitive judgment raise two questions: "What features of system 1 created the error?" and "Why was the error not detected and corrected by system 2?" (cf. Kahneman & Tversky, 1982). The first question is more basic, of course, but the second is also relevant and ought not be overlooked. Consider, for example, the paragraph that Tversky and Kahneman (1974; p. 3 in Kahneman, Slovic, & Tversky, 1982) used to introduced the notions of heuristic and bias:

> *The subjective assessment of probability resembles the subjective assessment of physical quantities such as distance or size. These judgments are all based on data of limited validity, which are processed according to heuristic rules. For example, the apparent distance of an object is determined in part by its clarity. The more sharply the object is seen, the closer it appears to be. This rule has some validity, because in any given scene the more distant objects are seen less sharply than nearer objects. However, the reliance on this rule leads to systematic errors in the estimation of distance. Specifically, distances are often overestimated when visibility is poor because the contours of objects are blurred. On the other hand, distances are often underestimated when visibility is good because the objects are seen sharply. Thus the reliance on clarity as an indication leads to common biases. Such biases are also found in intuitive judgments of probability.*

This statement was intended to extend Brunswik's (1943) analysis of the perception of distance to the domain of intuitive thinking and to provide a rationale for using biases to diagnose heuristics. However, the analysis of the effect of haze is flawed: It neglects the fact that an observer looking at a distant mountain possesses two relevant cues, not one. The first cue is the blur of the contours of the target mountain, which is positively correlated with its distance, when all else is equal. This cue should be given positive weight in a judgment of distance, and it is. The second relevant cue, which the observer can readily assess by looking around, is the ambient or general haziness.

In an optimal regression model for estimating distance, general haziness is a suppressor variable, which must be weighted negatively because it contributes to blur but is uncorrelated with distance. Contrary to the argument made in 1974, using blur as a cue does not inevitably lead to bias in the judgment of distance – the illusion could just as well be described as a failure to assign adequate negative weight to ambient haze. The effect of haziness on *impressions* of distance is a failing of system 1: The perceptual system is not designed to correct for this variable. The effect of haziness on *judgments* of distance is a separate failure of system 2. Although people are capable of consciously correcting their impressions of distance for the effects of ambient haze, they commonly fail to do so. A similar analysis applies to some of the judgmental biases we discuss later, in which errors and biases only occur when both systems fail.

In the following section, we present an attribute-substitution model of heuristic judgment, which assumes that difficult questions are often answered by substituting an answer to an easier one. This elaborates and extends earlier treatments of the topic (Kahneman & Tversky, 1982; Tversky & Kahneman, 1974, 1983). Following sections introduce a research design for studying attribute substitution, as well as discuss the controversy over the representativeness heuristic in the context of a dual-system view that we endorse. The final section situates representativeness within a broad family of prototype heuristics, in which properties of a prototypical exemplar dominate global judgments concerning an entire set.

Attribute Substitution

The early research on judgment heuristics was guided by a simple and general hypothesis: When confronted with a difficult question, people may answer an easier one instead and are often unaware of the substitution. A person who is asked "What proportion of long-distance relationships break up within a year?" may answer as if she had been asked "Do instances of failed long-distance relationships come readily to mind?" This would be an application of the availability heuristic. A professor who has heard a candidate's job talk and now considers the question "How likely is it that this candidate could be tenured in our department?" may answer the much easier question: "How impressive was the talk?". This would be an example of one form of the representativeness heuristic.

The heuristics and biases research program has focused primarily on representativeness and availability – two versatile attributes that are automatically computed and can serve as candidate answers to many different questions. It has also focused principally on thinking under uncertainty. However, the restriction to particular heuristics and to a specific context is largely arbitrary. Kahneman and Frederick (2002) argued that this process of *attribute substitution* is a general feature of heuristic judgment; that whenever the aspect of the judgmental object that one intends to judge (the *target attribute*) is less readily assessed than a related property that yields a plausible answer (the *heuristic attribute*), individuals may unwittingly substitute the simpler assessment. For an example, consider the well-known study by Strack, Martin, and Schwarz (1988) in which college students answered a survey that included these two questions: "How happy are you with your life in general?" and "How many dates did you have last month?" The correlation between the two questions was negligible when they occurred in the order shown, but rose to .66 if the dating question was asked first. We suggest that the question about dating frequency automatically evokes an evaluation of one's romantic satisfaction and that this evaluation lingers to become the heuristic attribute when the global happiness question is subsequently encountered.

To further illustrate the process of attribute substitution, consider a question in a study by Frederick and Nelson (2004): "If a sphere were dropped into a open cube, such that it just fit (the diameter of the sphere is the same as the interior width of the cube), what proportion of the volume of the cube would the sphere occupy?" The target attribute in this judgment (the volumetric relation between a cube and sphere) is simple enough to be understood but complicated enough to accommodate a wide range of estimates as plausible answers. Thus, if a relevant simpler computation or perceptual impression exists, respondents will have no strong basis for rejecting it as their "final answer." Frederick and Nelson (2004) proposed that the areal ratio of the respective cross-sections serves that function; that is, that respondents answer *as if* they were asked the simpler two-dimensional analog of this problem ("If a *circle* were drawn inside a *square*, what proportion of the *area* of the square does the circle occupy?"). As evidence, they noted that the mean estimate of the "sphere inside cube" problem (74%) is scarcely different from the mean estimate of the "circle inside square" problem (77%) and greatly exceeds the correct answer (52%) – a correct answer that most people, not surprisingly, are surprised by.

Biases

Whenever the heuristic attribute differs from the target attribute, the substitution of one for the other inevitably introduces systematic biases. In this treatment, we are mostly concerned with *weighting biases*, which arise when cues available to the judge are given either too much or too little weight. Criteria for determining optimal weights can be drawn from several sources. In the classic lens model, the optimal weights associated with different cues are the regression weights that optimize the prediction of an external criterion, such as physical distance or the grade point average that a college applicant will attain (Brunswik, 1943; Hammond, 1955). Our analysis of weighting biases applies to such cases, but it also extends to attributes for which no objective criterion is available, such as an individual's overall happiness or the probability that a particular patient will survive surgery. Normative standards for these attributes must be drawn from the constraints of ordinary language and are often imprecise. For example, the conventional interpretation of *overall happiness* does not specify how much weight ought to be given to various life domains. However, it certainly does require that substantial weight be given to every important domain of life and that no weight at all be given to the current weather or to the recent consumption of a cookie. Similar rules of common sense apply to judgments of probability. For example, the statement "John is more likely to survive a week than a month" is clearly true, and, thus, implies a rule that people would want their probability judgments to follow. Accordingly, neglect of duration in assessments of survival probabilities would be properly described as a weighting bias, even if there were no way to establish a normative probability for individual cases (Kahneman & Tversky, 1996).

For some judgmental tasks, information that could serve to supplement or correct the heuristic is not neglected or underweighted but simply lacking. If asked to judge the relative frequency of words beginning with K or R (Tversky & Kahneman, 1973) or to compare the population of a familiar foreign city with one that is unfamiliar (Gigerenzer & Goldstein, 1996), respondents have little recourse but to base their judgments on ease of retrieval or recognition. The necessary reliance on these heuristic attributes renders such judgments susceptible to biasing factors (e.g., the amount of media coverage). However, unlike weighting biases, such biases of insufficient information cannot be described as errors of judgment because there is no way to avoid them.

Accessibility and Substitution

The intent to judge a target attribute initiates a search for a reasonable value. Sometimes this search ends quickly because the required value can be read from a stored memory (e.g., the answer to the question "How tall are you?") or a current experience (e.g., the answer to the question "How much do you like this cake?"). For other judgments, however, the target attribute does not readily come to mind, but the search for it evokes other attributes that are conceptually and associatively related. For example, a question about overall happiness may retrieve the answer to a related question about satisfaction with a particular aspect of life upon which one is currently reflecting.

We adopt the term *accessibility* to refer to the ease (or effort) with which particular mental contents come to mind (see, e.g., Higgins, 1996; Tulving & Pearlstone, 1966). The question of why thoughts become accessible – why particular ideas come to mind at particular times – has a long history in psychology and encompasses notions of stimulus salience, associative activation, selective attention, specific training, and priming. In the present usage, *accessibility* is determined jointly by the characteristics of the cognitive mechanisms that produce it and by the characteristics of the stimuli and events that evoke it, and it may refer to different aspects and elements of a situation, different objects in a scene, or different attributes of an object.

Attribute substitution occurs when a relatively inaccessible target attribute is assessed by mapping a relatively accessible and related heuristic attribute onto the target scale. Some attributes are permanent candidates for the heuristic role because they are routinely evaluated as part of perception and comprehension and therefore always accessible (Tversky & Kahneman, 1983). These *natural assessments* include physical properties such as size and distance and more abstract properties such as similarity (e.g., Tversky & Kahneman, 1983; see Goldstone & Son, Chap. 2), cognitive fluency in perception and memory (e.g., Jacoby & Dallas, 1991; Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973), causal propensity (Heider, 1944; Kahneman & Varey, 1990; Michotte, 1963), surprisingness (Kahneman & Miller, 1986), mood (Schwarz & Clore, 1983), and affective valence (e.g., Bargh, 1997; Cacioppo, Priester, & Berntson, 1993; Kahneman, Ritov, & Schkade, 1999; Slovic et al., 2002; Zajonc, 1980, 1997).

Because affective valence is a natural assessment, it is a candidate for attribute substitution in a wide variety of affect-laden judgments. Indeed, the evidence suggests that a list of major general-purpose heuristics should include an *affect heuristic* (Slovic et al., 2002). Slovic and colleagues (2002) show that a basic affective reaction governs a wide variety of more complex evaluations such as the cost–benefit ratio of various technologies, the safe level of chemicals, or even the predicted economic performance of various industries. In the same vein, Kahneman and Ritov (1994) and Kahneman, Ritov, and Schkade (1999) proposed that an automatic affective valuation is the principal determinant of willingness to pay for public goods, and Kahneman, Schkade, and Sunstein (1998) interpreted jurors' assessments of punitive awards as a mapping of outrage onto a dollar scale of punishments.

Attributes that are not naturally assessed can become accessible if they have been recently evoked or primed (see, e.g., Bargh et al., 1986; Higgins & Brendl, 1995). The effect of temporary accessibility is illustrated by the "romantic satisfaction heuristic" for judging happiness. The mechanism of attribute substitution is the same, however, whether the heuristic attribute is chronically or temporarily accessible.

There is sometimes more than one candidate for the role of heuristic attribute. For an example that we borrow from Anderson (1991), consider the question "Are more deaths caused by rattlesnakes or bees?" A respondent who has recently read about someone who died from a snakebite or bee sting may use the relative availability of instances of the two categories as a heuristic. If no instances come to mind, that person might consult his or her impressions of the "dangerousness" of the typical snake or bee, an application of representativeness. Indeed, it is possible that the question initiates both a search for instances and an assessment of dangerousness, and that a contest of accessibility determines the role of the two heuristics in the final response. As Anderson observed, it is not always possible to determine a priori which heuristic will govern the response to a particular problem.

The original list of heuristics (Tversky & Kahneman, 1974) also included an "anchoring heuristic." An anchoring effect, however, does not involve the substitution of a heuristic attribute for a target attribute: It is due to the temporary salience of a particular value of the target attribute. However, anchoring and attribute substitution are both instances of a broader family of *accessibility effects* (Kahneman, 2003). In attribute substitution, a highly accessible attribute controls the evaluation of a less accessible one. In anchoring, a highly accessible *value* of the target attribute dominates its judgment. This conception is compatible with more recent theoretical treatments of anchoring (see, e.g., Chapman & Johnson, 1994, 2002; Mussweiler & Strack 1999; Strack & Mussweiler, 1997).

Cross-Dimensional Mapping

The process of attribute substitution involves the mapping of the heuristic attribute of the judgment object onto the scale of the target attribute. Our notion of cross-dimensional mapping extends Stevens' (1975) concept of cross-modality matching. Stevens postulated that intensive attributes (e.g., brightness, loudness, the severity of crimes) can be mapped onto a common scale of sensory strength, allowing direct matching of intensity across modalities – permitting, for example, respondents to match the loudness of sounds to the severity of crimes. Our conception allows other ways of comparing values across dimensions, such as matching relative positions (e.g., percentiles) in the frequency distributions or ranges of different attributes (Parducci, 1965). An impression of a student's position in the distribution of aptitude may be mapped directly onto a corresponding position in the distribution of academic achievement and then translated into a letter grade. Note that cross-dimensional matching is inherently nonregressive: A judgment or prediction is just as extreme as the impression mapped onto it. Ganzach and Krantz (1990) applied the term "univariate matching" to a closely related notion.

Cross-dimensional mapping presents special problems when the scale of the target attribute has no upper bound. Kahneman, Ritov, and Schkade (1999) discussed two situations in which an attitude (or affective valuation) is mapped onto an unbounded scale of dollars: when respondents in surveys are required to indicate how much money they would contribute for a cause, and when jurors are required to specify an amount of punitive damages against a negligent firm. The mapping of attitudes onto dollars is a variant of direct scaling in psychophysics, where respondents assign numbers to indicate the intensity of sensations (Stevens, 1975). The normal practice of direct scaling is for the experimenter to provide a *modulus* – a specified number that is to be associated with a standard stimulus. For example, respondents may be asked to assign the number 10 to the loudness of a standard sound and judge the loudness of other sounds relative to that standard. Stevens (1975) observed that when the experimenter fails to provide a modulus, respondents spontaneously adopt one. However, different respondents may pick moduli that differ greatly (sometimes varying by a factor of 100 or more); thus, the variability in judgments of particular stimuli is dominated by arbitrary individual differences in the choice of modulus. A similar analysis applies to situations in which respondents are required to use the dollar scale to express affection for a species or outrage toward a defendant. Just as Stevens' observers had no principled way to assign a number to a moderately loud sound, survey participants and jurors have no principled way to scale affection or outrage into dollars. The analogy of scaling without a modulus has been used to explain the notorious variability of dollar responses in surveys of willingness to pay and in jury awards (Kahneman, Ritov, & Schkade, 1999; Kahneman, Schkade, & Sunstein, 1998).

System 2: The Supervision of Intuitive Judgments

Our model assumes that an intuitive judgment is expressed overtly only if it is endorsed by system 2. The Stroop task illustrates this two-system structure. Observers who are instructed to report the color in which words are printed tend to stumble when the word is the name of another color (e.g., the word BLUE printed in green ink). The difficulty arises because the word is automatically read, and activates a response ("blue" in this case) that competes with the required response ("green"). Errors are rare in the Stroop test, indicating generally successful monitoring and control of the overt response, but the conflict produces delays and hesitations. The successful suppression of erroneous responses is effortful, and its efficacy is reduced by stress and distraction.

Gilbert (1989) described a correction model in which initial impulses are often wrong and normally overridden. He argued that people initially believe whatever they are told (e.g., "Whitefish love grapes") and that it takes some time and mental effort to "unbelieve" such dubious statements. Here again, cognitive load disrupts the controlling operations of system 2, increasing the rate of errors and revealing aspects of intuitive thinking that are normally suppressed. In an ingenious extension of this approach, Bodenhausen (1990) exploited natural temporal variability in alertness. He found that "morning people" were substantially more susceptible to a judgment bias (the conjunction fallacy) in the evening and that "evening people" were more likely to commit the fallacy in the morning.

Because system 2 is relatively slow, its operations can be disrupted by time pressure. Finucane et al. (2000) reported a study in which respondents judged the risks and benefits of various products and technologies (e.g., nuclear power, chemical plants, cellular phones). When participants were forced to respond within 5 seconds, the correlations between their judgments of risks and their judgments of benefits were strongly negative. The negative correlations were much weaker (although still pronounced) when respondents were given more time to ponder a response. When time is short, the same affective evaluation apparently serves as a heuristic attribute for assessments of both benefits and risks. Respondents can move beyond this simple strategy, but they need more than 5 seconds to do so. As this example illustrates, judgment by heuristic often yields simplistic assessments, which system 2 sometimes corrects by bringing additional considerations to bear.

Attribute substitution can be prevented by alerting respondents to the possibility that their judgment could be contaminated by an irrelevant variable. For example, although sunny or rainy weather typically affects reports of well-being, Schwarz and Clore (1983) found that weather has no effect if respondents are asked about the weather just before answering the well-being question. Apparently, this question reminds respondents that their current mood (a candidate heuristic attribute) is influenced by a factor (current weather) that is irrelevant to the requested target attribute (overall well-being). Schwarz (1996) also found that asking people to describe their satisfaction with some particular domain of life *reduces* the weight this domain receives in a subsequent judgment of overall well being. As these examples illustrate, although priming typically increases the weight of that variable on judgment (a system 1 effect), this does not occur if the prime is a sufficiently explicit reminder that brings the self-critical operations of system 2 into play.

We suspect that system 2 endorsements of intuitive judgments are granted quite casually under normal circumstances. Consider the puzzle "A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?" Almost everyone we ask reports an initial tendency to answer "10 cents" because the sum $1.10 separates naturally into $1 and 10 cents, and 10 cents is about the right magnitude. Many people yield to this immediate impulse. Even among undergraduates at elite institutions, about half get this problem wrong when it is included in a short IQ test (Frederick, 2004). The critical feature of this problem is that anyone who reports 10 cents has obviously not taken the trouble to check his or her answer. The surprisingly high rate of errors in this easy problem illustrates how lightly system 2 monitors the output of system 1: People are often content to trust a plausible judgment that quickly comes to mind. (The correct answer, by the way, is 5 cents.)

The bat and ball problem elicits many errors, although it is not really difficult and certainly not ambiguous. A moral of this example is that people often make quick intuitive judgments to which they are not deeply committed. A related moral is that we should be suspicious of analyses that explain apparent errors by attributing to respondents a bizarre interpretation of the question. Consider someone who answers a question about happiness by reporting her satisfaction with her romantic life. The respondent is surely not committed to the absurdly narrow interpretation of happiness that her response seemingly implies. More likely, at the time of answering, she thinks that she *is* reporting happiness: A judgment comes quickly to mind and is not obviously mistaken – end of story. Similarly, we propose that respondents who judge probability by representativeness do not seriously believe that the questions "How likely is X to be a Y?" and "How much does X resemble the stereotype of Y?" are synonymous. People who make a casual intuitive judgment normally know little about how their judgment came about and know even less about its logical entailments. Attempts to reconstruct the meaning of intuitive judgments by interviewing respondents (see, e.g., Hertwig & Gigerenzer, 1999) are therefore unlikely to succeed because such probes require better introspective access and more coherent beliefs than people normally muster.

Identifying a Heuristic

Hypotheses about judgment heuristics have most often been studied by examining weighting biases and deviations from normative rules. However, the hypothesis that one attribute is substituted for another in a judgment task – for example, representativeness for probability – can also be tested more directly. In the *heuristic elicitation* design, one group of respondents provides judgments of a target attribute for a set of objects and another group evaluates the hypothesized heuristic attribute for the same objects. The substitution hypothesis implies that the judgments of the two groups, when expressed in comparable units (e.g., percentiles), will be identical. This section examines several applications of heuristic elicitation.

Eliciting Representativeness

Figure 12.1 displays the results of two experiments in which a measure of representativeness was elicited. These results were published long ago, but we repeat them here because they still provide the most direct evidence for both attribute substitution and the representativeness heuristic. For a more recent application of a similar design, see Bar-Hillel and Neter (1993).

The object of judgment in the study from which Figure 12.1(a) is drawn (Kahneman & Tversky, 1973; p. 127 in Kahneman, Slovic, & Tversky, 1982) was the following description of a fictitious graduate student, which was shown along with a list of nine fields of graduate specialization:

> *Tom W. is of high intelligence, although lacking in true creativity. He has a need for order and clarity and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense.*

Participants in a *representativeness* group ranked the nine fields of specialization by the degree to which Tom W. "resembles a typical graduate student." Participants in the *probability* group ranked the nine fields according to the likelihood of Tom W.'s specializing in each. Figure 12.1(a) plots the mean judgments of the two groups. The correlation between representativeness and

P1: IKB-IRK/KAB P2: IKB-IYP/KAB QC: IYP/JZO T1: KOD
0521824176c12.xml CB798B/Holyoak 0 521 82417 6 February 2, 2005 0:41

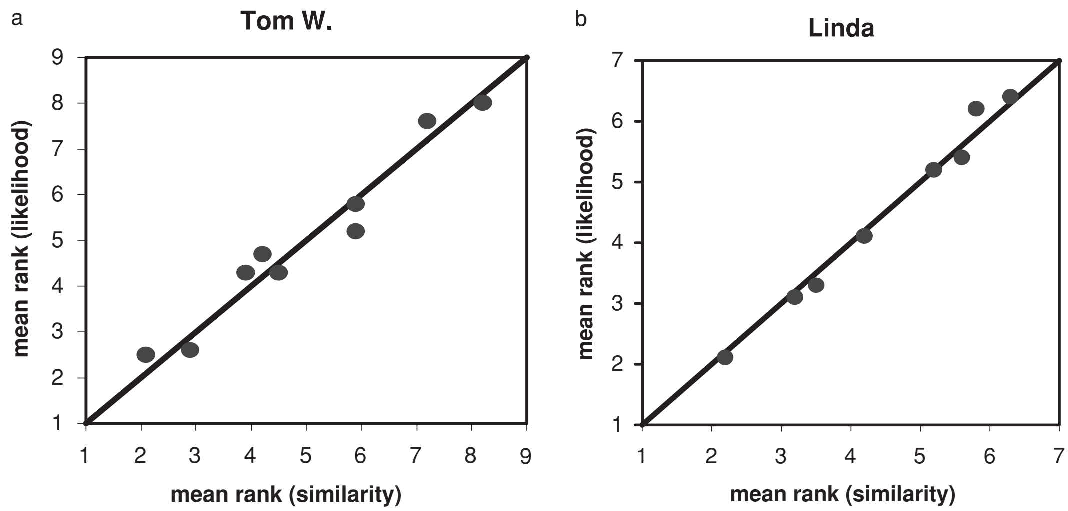

Figure 12.1. (a) Plot of average ranks for nine outcomes for Tom W. ranked by probability and by similarity to stereotypes of graduate students in various fields. (b) Plot of average ranks for eight outcomes for Linda ranked by probability and by representativeness.

probability is nearly perfect (.97). No stronger support for attribute-substitution could be imagined. However, interpreting representativeness as the heuristic attribute in these judgments does require two additional plausible assumptions – that representativeness is more accessible than probability, and that there is no third attribute that could explain both judgments.

The Tom W. study was also intended to examine the effect of the base rates of outcomes on categorical prediction. For that purpose, respondents in a third group estimated the proportion of graduate students enrolled in each of the nine fields. By design, some outcomes were defined quite broadly, whereas others were defined more narrowly. As intended, estimates of base rates varied markedly across fields, ranging from 3% for Library Science to 20% for Humanities and Education. Also by design, the description of Tom W. included characteristics (e.g., introversion) that were intended to make him fit the stereotypes of the smaller fields (library science, computer science) better than the larger fields (humanities and social sciences).[1] As intended, the correlation between the average judgments of representativeness and of base rates was strongly negative (−.65).

The logic of probabilistic prediction in this task suggests that the ranking of outcomes by their probabilities should be intermediate between their rankings by representativeness and by base rate frequencies. Indeed, if the personality description is taken to be a poor source of information, probability judgments should stay quite close to the base rates. The description of Tom W. was designed to allow considerable scope for judgments of probability to diverge from judgments of representativeness, as this logic requires. Figure 12.1 (a) shows no such divergence. Thus, the results of the Tom W. study simultaneously demonstrate the substitution of representativeness for probability and the neglect of known (but not explicitly mentioned) base rates.

Figure 12.1 (b) is drawn from an early study of the Linda problem, the best-known and most controversial example in the representativeness literature (Tversky & Kahneman, 1982) in which a woman named Linda was described as follows:

> *Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations.*

As in the Tom W. study, separate groups of respondents were asked to rank a set of eight outcomes by representativeness and probability. The results are shown in Figure 12.1(b). Again the correlation between these rankings was almost perfect (.99).[1]

Six of the eight outcomes that subjects were asked to rank were fillers (e.g., elementary school teacher, psychiatric social worker). The two critical outcomes were #6 (bank teller) and the so-called conjunction item #8 (bank teller and active in the feminist movement). Most subjects ranked the conjunction higher than its constituent, both in representativeness (85%) and probability (89%). The observed ranking of the two items is quite reasonable for judgments of similarity, but not for probability: Linda may resemble a feminist bank teller more than she resembles a bank teller, but she cannot be more likely to be a feminist bank teller than to be a bank teller. In this problem, reliance on representativeness yields probability judgments that violate a basic logical rule. As in the Tom W. study, the results make two points: They support the hypothesis of attribute substitution and also illustrate a predictable judgment error.

The Representativeness Controversy

The experiments summarized in Figure 12.1 provided direct evidence for the representativeness heuristic and two concomitant biases: neglect of base rates and conjunction errors. In the terminology introduced by Tversky and Kahneman (1983), the design of these experiments was "subtle": Adequate information was available for participants to avoid the error, but no effort was made to call their attention to that information. For example, participants in the Tom W. experiment had general knowledge of the relative base rates of the various fields of specialization, but these base rates were not explicitly mentioned in the problem. Similarly, both critical items in the Linda experiment were included in the list of outcomes, but they were separated by a filler so respondents would not feel compelled to compare them. In the anthropomorphic language used here, system 2 was given a chance to correct the judgment but was not prompted to do so.

In view of the confusing controversy that followed, it is perhaps unfortunate that the articles documenting base rate neglect and conjunction errors did not stop with subtle tests. Each article also contained an experimental flourish – a demonstration in which the error occurred in spite of a manipulation that called participants' attention to the critical variable. The engineer–lawyer problem (Kahneman & Tversky, 1973) included special instructions to ensure that respondents would notice the base rates of the outcomes. The brief personality descriptions shown to respondents were reported to have been drawn from a set containing descriptions of 30 lawyers and 70 engineers (or vice versa), and respondents were asked "What is the probability that this description belongs to one of the 30 lawyers in the sample of 100?" To the authors' surprise, base rates were largely neglected in the responses, despite their salience in the instructions. Similarly, the authors were later shocked to discover that more than 80% of undergraduates committed a conjunction error even when asked point blank whether Linda was more likely to be "a bank teller" or "a bank teller who is active in the feminist movement" (Tversky & Kahneman, 1983). The novelty of these additional direct or "transparent" tests was the finding that respondents continued to show the biases associated with representativeness even in the presence of strong cues pointing to the normative response. The errors that people make in transparent judgment problems are analogous to observers' failure to allow for ambient haze in estimating distances: A correct response is within reach, but not chosen, and the failure involves an unexpected weakness of the corrective operations of system 2.

Discussions of the heuristics and biases approach have focused almost exclusively on the direct conjunction fallacy and on the engineer–lawyer problems. These are also the only studies that have been extensively replicated with varying parameters. The amount of critical attention is remarkable because the studies were not, in fact, essential to the authors' central claim. In terms of the present treatment, the claim was that intuitive prediction is an operation of system 1, which is susceptible to both base rate neglect and conjunction fallacies. There was no intent to deny the possibility of system 2 interventions that would modify or override intuitive predictions. Thus, the articles in which these studies appeared would have been substantially the same, although far less provocative, if respondents had overcome base rate neglect and conjunction errors in transparent tests.

To appreciate why the strong forms of base rate neglect and of the conjunction fallacy sparked so much controversy, it is useful to distinguish two conceptions of human rationality (Kahneman, 2000b). *Coherence rationality* is the strict conception that requires the agent's entire system of beliefs and preferences to be internally consistent and immune to effects of framing and context. For example, an individual's probability *p* ("Linda is a bank teller") should be the sum of the probabilities *p* ("Linda is a bank teller and a feminist"), and *p* ("Linda is a bank teller and not a feminist"). A subtle test of coherence rationality could be conducted by asking individuals to assess these three probabilities on separate occasions under circumstances that minimize recall. Coherence can also be tested in a between-groups design. If random assignment is assumed, the sum of the average probabilities assigned to the two component events should equal the average judged probability of "Linda is a bank teller." If this prediction fails, then at least some individuals are incoherent. Demonstrations of incoherence present a significant challenge to important models of decision theory and economics, which attribute to agents a very strict form of rationality (Tversky & Kahneman, 1986). Failures of perfect coherence are less provocative to psychologists, who have a more realistic view of human capabilities.

A more lenient concept, *reasoning rationality*, only requires an ability to reason correctly about the information currently at hand without demanding perfect consistency among beliefs that are not simultaneously evoked. The best known violation of reasoning rationality is the famous "four card" problem (Wason, 1960). The failure of intelligent adults to reason their way through this problem is surprising because the problem is "easy" in the sense of being easily understood once explained. What everyone learns, when first told that intelligent people fail to solve the four-card problem, is that one's expectations about human reasoning abilities had not been adequately calibrated. There is, of course, no well-defined metric of reasoning rationality, but whatever metric one uses, the Wason problem calls for a downward adjustment. The surprising results of the Linda and engineer–lawyer problems led Tversky and Kahneman to a similar realization: The reasoning of their subjects was less proficient than they had anticipated. Many readers of the work shared this conclusion, but many others strongly resisted it.

The implicit challenge to reasoning rationality was met by numerous attempts to dismiss the findings of the engineer–lawyer and the Linda studies as artifacts of ambiguous language, confusing instructions, conversational norms, or inappropriate normative standards. Doubts have been raised about the proper interpretation of almost every word in the conjunction problem, including "bank teller," "probability," and even "and" (see, e.g., Dulany & Hilton, 1991; Hilton & Slugoski, 2001). These claims are not discussed in detail here. We suspect that most of them have some validity and that they identified mechanisms that may have made the results in the engineer–lawyer and Linda studies exceptionally strong. However, we note a significant weakness shared by all these critical discussions: They provide no explanation of the essentially perfect consistency of the judgments observed in direct tests of the conjunction rule and in three other types of experiments: subtle comparisons, between-Ss comparisons, and most important, judgments of representativeness (see also Bar-Hillel & Neter, 1993). Interpretations of the conjunction fallacy as an artifact implicitly dismiss the results of Figure 12.1(b) as a coincidence (for an exception, see Ayton, 1998). The story of the engineer-lawyer problem is similar. Here again, multiple demonstrations in which base rate information was used (see Koehler, 1996, for a review) invited the inference that there is no general problem of base rate neglect. Again, the data of prediction by representativeness in Figure 12.1(a) (and related results reported by Kahneman & Tversky, 1973) were ignored.

The demonstrations that under some conditions people avoid the conjunction fallacy in direct tests, or use explicit base rate information, led some scholars to the blanket conclusion that judgment biases are artificial and fragile and that there is no need for judgment heuristics to explain them. This position was promoted most vigorously by Gigerenzer (1991). Kahneman and Tversky (1996) argued in response that the heuristics and biases position does not preclude the possibility of people's performing flawlessly in particular variants of the Linda and the engineer–lawyer problems. Because laypeople readily acknowledge the validity of the conjunction rule and the relevance of base rate information, the fact that they sometimes obey these principles is neither a surprise nor an argument against the role of representativeness in routine intuitive prediction. However, the study of conditions under which errors are avoided can help us understand the capabilities and limitations of system 2. We develop this argument further in the next section.

Making Biases Disappear: A Task for System 2

Much has been learned over the years about variables and experimental procedures that reduce or eliminate the biases associated with representativeness. We next discuss conditions under which errors of intuition are successfully overcome and some circumstances under which intuitions may not be evoked at all.

STATISTICAL SOPHISTICATION

The performance of statistically sophisticated groups of respondents in different versions of the Linda problem illustrates the effects of both expertise and research design (Tversky & Kahneman, 1983). Statistical expertise provided no advantage in the eight-item version in which the critical items were separated by a filler and were presumably considered separately. In the two-item version, in contrast, respondents were effectively compelled to compare "bank teller" with "bank teller and is active in the feminist movement." The incidence of conjunction errors remained essentially unchanged among the statistically naive in this condition but dropped dramatically for the statistically sophisticated. Most of the experts followed logic rather than intuition when they recognized that one of the categories contained the other. In the absence of a prompt to compare the items, however, the statistically sophisticated made their predictions in the same way as everyone else does – by representativeness. As Stephen Jay Gould (1991, p. 469) noted, knowledge of the truth does not dislodge the feeling that Linda is a feminist bank teller: "I know [the right answer], yet a little homunculus in my head continues to jump up and down, shouting at me – 'but she can't just be a bank teller; read the description.'"

INTELLIGENCE

Stanovich (1999) and Stanovich and West (2002) observed a generally negative correlation between conventional measures of intelligence and susceptibility to judgment biases. They used transparent versions of the problems, which include adequate cues to the correct answer and therefore provide a test of reasoning rationality. Not surprisingly, intelligent people are more likely to possess the relevant logical rules and also to recognize the applicability of these rules in particular situations. In the terms of the present analysis, high-IQ respondents benefit from relatively efficient system 2 operations that enable them to overcome erroneous intuitions when adequate information is available. (However, when a problem is too difficult for everyone, the correlation may reverse because the more intelligent respondents are more likely to agree on a plausible error than to respond randomly, as discussed in Kahneman, 2000b.)

FREQUENCY FORMAT

Relative frequencies (e.g., 1 in 10) are more vividly represented and more easily understood than equivalent probabilities (.10) or percentages (10%). For example, the emotional impact of statements of risk is enhanced by the frequency format: "1 person in 1000 will die" is more frightening than a probability of .001 (Slovic et al., 2002). The frequency representation also makes it easier to visualize partitions of sets and detect that one set is contained in another. As a consequence, the conjunction fallacy is generally avoided in direct tests in which the frequency format makes it easy to recognize that feminist bank tellers are a subset of bank tellers (Gigerenzer & Hoffrage, 1995; Tversky & Kahneman, 1983). For similar reasons, some base rate problems are more easily solved when couched in frequencies than in probabilities or percentages (Cosmides & Tooby, 1996). However, there is little support for the more general claims about the evolutionary adaptation of the mind to deal with frequencies (Evans et al., 2000). Furthermore, the ranking of outcomes by predicted relative frequency is very similar to the ranking of the same outcomes by representativeness (Mellers, Hertwig, & Kahneman, 2001). We conclude that the frequency format affects the corrective operations of system 2, not the intuitive operations of system 1. The language of frequencies improves respondents' ability to impose the logic of set inclusion on their considered judgments but does not reduce the role of representativeness in their intuitions.

MANIPULATIONS OF ATTENTION

The weight of neglected variables can be increased by drawing attention to them, and experimenters have devised many ingenious ways to do so. Schwarz et al. (1991) found that respondents pay more attention to base rate information when they are instructed to think as statisticians rather than clinical psychologists. Krosnick, Li, and Lehman (1990) exploited conversational conventions about the sequencing of information and confirmed that the impact of base rate information was enhanced by presenting that information *after* the personality description rather than before it. Attention to the base rate is also enhanced when participants observe the drawing of descriptions from an urn (Gigerenzer, Hell, & Blank, 1988) perhaps because watching the drawing induces conscious expectations that reflect the known proportions of possible outcomes. The conjunction fallacy can also be reduced or eliminated by manipulations that increase the accessibility of the relevant rule, including some linguistic variations (Macchi, 1995), and practice with logical problems (Agnoli, 1991; Agnoli & Krantz, 1989).

The interpretation of these attentional effects is straightforward. We assume most participants in judgment studies know, at least vaguely, that the base rate is relevant and that the conjunction rule is valid (Kahneman & Tversky, 1982). Whether they apply this knowledge to override an intuitive judgment depends on their cognitive skills (education, intelligence) and on formulations that make the applicability of a rule apparent (frequency format) or a relevant factor more salient (manipulations of attention). We assume intuitions are less sensitive to these factors and that the appearance or disappearance of biases mainly reflects variations in the efficacy of corrective operations. This conclusion would be circular, of course, if the corrective operations were both inferred from the observation of correct performance and used to explain that performance. Fortunately, the circularity can be avoided because the role of system 2 can be verified – for example, by using manipulations of time pressure, cognitive load, or mood to interfere with its operations.

WITHIN-SUBJECTS FACTORIAL DESIGNS

The relative virtues of between-subjects and within-subject designs in studies of judgment are a highly contentious issue. Factorial designs have their dismissive critics (e.g., Poulton, 1989) and their vigorous defenders (e.g., Birnbaum, 1999). We do not attempt to adjudicate this controversy here. Our narrower point is that between-subjects designs are more appropriate for the study of heuristics of judgment. The following arguments favor this conclusion:

- Factorial designs are transparent. Participants are likely to identify the variables that are manipulated, especially if there are many trials and especially in a fully factorial design in which the same stimulus attributes are repeated in varying combinations. The message that the design conveys to the participants is that the experimenter expects to find effects of every factor that is manipulated (Bar-Hillel & Fischhoff, 1981; Schwarz, 1996).
- Studies that apply a factorial design to judgment tasks commonly involve schematic and impoverished stimuli. The tasks are also highly repetitive. These features encourage participants to adopt simple mechanical rules that will allow them to respond quickly without forming an individuated impression of each stimulus. For example, Ordóñez and Benson (1997) required respondents to judge the attractiveness of gambles on a 100-point scale. They found that under time pressure many respondents computed or estimated the expected values of the gambles and used the results as attractiveness ratings (e.g., a rating of 15 for a 52% chance to win $31.50).
- Factorial designs often yield judgments that are linear combinations of the manipulated variables. This is a central conclusion of a massive research effort conducted by Anderson (1996), who observed that people often average or add where they should multiply.

In summary, the factorial design is not appropriate for testing hypotheses about biases of neglect because it effectively guarantees that no manipulated factor is neglected. Figure 12.2 illustrates this claim by several examples of an *additive extension effect* that we discuss further in the next section. The experiments summarized in the different panels share three important features: (1) In each case, the quantitative variable plotted on the abscissa was completely neglected in similar experiments conducted in a between-subjects or subtle design; (2) in each case, the quantitative variable combines additively with other information; (3) in each case, a compelling normative argument can be made for a quasimultiplicative rule in which the lines shown in Figure 12.2 should fan out. For example, Figure 12.2(c) presents a study of categorical prediction (Novemsky & Kronzon, 1999) in which respondent 5 judged the relative likelihood that a person was a member of one occupation rather than another (e.g., computer programmer vs. flight attendant) on the basis of short personality sketches (e.g., "shy, serious, organized, and sarcastic") and one of three specified base rates (10%, 50%, or 90%). Representativeness and base rate were varied factorially within subjects. The effect of base rate is clearly significant in this design (see also Birnbaum & Mellers, 1983). Furthermore, the effects of representativeness and base rate are strictly additive. As Anderson (1996) argued, averaging (a special case of additive combination) is the most obvious way to combine the effects of two variables that are recognized as relevant (e.g., "She looks like a bank teller, but the base-rate is low."). Additivity is not normatively appropriate in this case – any Bayes-like combination would produce curves that initially fan out from the origin and converge again at high values. Similar considerations apply to the other three panels of Figure 12.2 discussed later. Between-subjects and factorial designs often yield different results in studies of intuitive judgment. Why should we

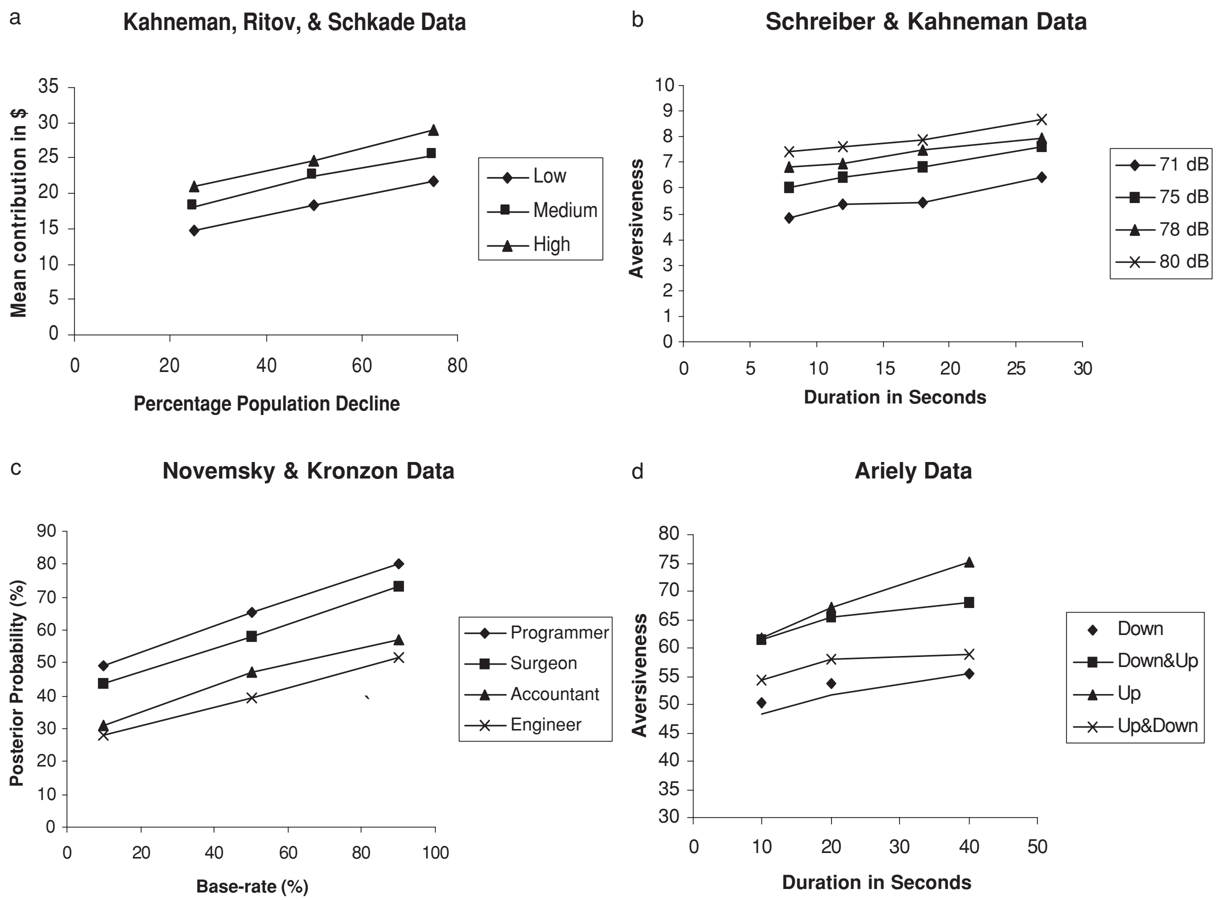

Figure 12.2. (a) Willingness to pay to restore damage to species that differ in popularity as a function of the damage they have suffered (from Kahneman, Ritov, & Schkade 2000); (b) global evaluations of aversive sounds of different loudness as a function of duration for subjects selected for their high sensitivity to duration (from Schreiber & Kahneman, 2000); (c) ratings of probability for predictions that differ in representativeness as a function of base rate frequency (from Novemsky & Kronzon, 1999); (d) global evaluations of episodes of painful pressure that differ in temporal profile as a function of duration (Ariely, 1998).

believe one design rather than the other? The main argument against the factorial design is its poor ecological validity. Encountering multiple judgment objects in rapid succession in a rigidly controlled structure is unique to the laboratory, and the solutions that they evoke are not likely to be typical. Direct comparisons among concepts that differ in only one variable – such as bank teller and feminist bank tellers – also provide a powerful hint and a highly unusual opportunity to overcome intuitions. The between-subjects design, in contrast, mimics the haphazard encounters in which most judgments are made and is more likely to evoke the casually intuitive mode of judgment that governs much of mental life in routine situations (e.g., Langer, 1978).

Prototype Heuristics and the Neglect of Extension

In this section, we offer a common account of three superficially dissimilar judgmental tasks: (1) categorical prediction (e.g., "*In a set of 30 lawyers and 70 engineers, what is the* *probability that someone described as 'charming, talkative, clever, and cynical' is one of the lawyers?*"); (2) summary evaluations of past events (e.g., "*Overall, how aversive was it to be exposed for 30 minutes to your neighbor's car alarm?*"); and (3) economic valuations of public goods (e.g., "*What is the most you would be willing to pay to prevent 200,000 migrating birds from drowning in uncovered oil ponds?*"). We propose that a generalization of the representativeness heuristic accounts for the remarkably similar biases that are observed in these diverse tasks.

The original analysis of categorical prediction by representativeness (Kahneman & Tversky 1973; Tversky & Kahneman, 1983) invoked two assumptions in which the word "representative" was used in different ways: (1) A prototype (a *representative* exemplar) is used to represent categories (e.g., bank tellers) in the prediction task, and (2) the probability that the individual belongs to a category is judged by the degree to which the individual resembles (is *representative* of) the category stereotype. Thus, categorical prediction by representativeness involves two separate acts of substitution – the substitution of a representative exemplar for a category and the substitution of the heuristic attribute of representativeness for the target attribute of probability. Perhaps because they share a label, the two processes have not been distinguished in discussions of the representativeness heuristic. We separate them here by describing *prototype heuristics* in which a prototype is substituted for its category, but in which representativeness is not necessarily the heuristic attribute.

The target attributes to which prototype heuristics are applied are *extensional*. An extensional attribute pertains to an aggregated property of a set or category for which an extension is specified – the probability that a set of 30 lawyers includes Jack, the overall unpleasantness of a set of moments of hearing a neighbor's car alarm, and the personal dollar value of saving a certain number of birds from drowning in oil ponds. Normative judgments of extensional attributes are governed by a general principle of conditional adding, which dictates that each element of the set adds to the overall judgment an amount that depends on the elements already included. In simple cases, conditional adding is just regular adding – the total weight of a collection of chairs is the sum of their individual weights. In other cases, each element of the set contributes to the overall judgment, but the combination rule is not simple addition and is most typically subadditive. For example, the economic value of protecting *X* birds should be increasing in *X*, but the value of saving 2000 birds is for most people less than twice as large as the value of saving 1000 birds.

The logic of categorical prediction entails that the probability of membership in a category should vary with its relative size, or base rate. In prediction by representativeness, however, the representation of outcomes by prototypical exemplars effectively discards base rates because the prototype of a category (e.g., lawyers) contains no information about the size of its membership. Next, we show that phenomena analogous to the neglect of *base rate* are observed in other prototype heuristics: The monetary value attached to a public good is often insensitive to its *scope*, and the global evaluation of a temporally extended experience is often insensitive to its *duration*. These various instantiations of *extension neglect* (neglect of base rates, scope, and duration) have been discussed in separate literatures, but all can be explained by the two-part process that defines prototype heuristics: (1) A category is represented by a prototypical exemplar, and (2) a (nonextensional) property of the prototype is then used as a heuristic attribute to evaluate an extensional target attribute of the category. As might be expected from the earlier discussion of base rate neglect, extension neglect in all its forms is most likely to be observed in between-subjects experiments. Within-subject factorial designs consistently yield the *additive extension effect* illustrated in Figure 12.2.

Scope Neglect in Willingness to Pay

The contingent valuation method (CVM) was developed by resource economists (see Mitchell & Carson, 1989) as a tool for assessing the value of public goods for purposes of litigation or cost–benefit analysis. Participants in contingent valuation (CV) surveys are asked to indicate their willingness to pay (WTP) for specified public goods, and their responses are used to estimate the total amount that the community would pay to obtain these goods. The economists who design contingent valuation surveys interpret WTP as a valid measure of economic value and assume that statements of WTP conform to the extensional logic of consumer theory. The relevant logic has been described by a critic of CVM (Diamond, 1996), who illustrates the conditional adding rule by the following example: In the absence of income effects, WTP for saving X birds should equal WTP for saving $(X - k)$ birds, plus WTP to save k birds, where the last value is contingent on the costless prior provision of safety for $(X - k)$ birds.

Strict adherence to Bayes' rule may be an excessively demanding standard for intuitive predictions; similarly, it would be too much to ask for WTP responses that strictly conform to the "add-up rule." In both cases, however, it seems reasonable to expect *some* sensitivity to extension – to the base rate of outcomes in categorical prediction and to the scope of the good in WTP. In fact, several studies have documented nearly complete neglect of scope in CV surveys. The best-known demonstration of scope neglect is an experiment by Desvouges et al. (1993), who used the scenario of migratory birds that drown in oil ponds. The number of birds said to die each year was varied across groups. The WTP responses were completely insensitive to this variable; the mean WTPs for saving 2000, 20,000, or 200,000 birds were \$80, \$78, and \$88, respectively.

A straightforward interpretation of this result involves the two acts of substitution that characterize prototype heuristics. The deaths of numerous birds are first represented by a prototypical instance – perhaps an image of a bird soaked in oil and drowning. The prototype automatically evokes an affective response, and the intensity of that emotion is then mapped onto the dollar scale – substituting the readily accessible heuristic attribute of affective intensity for the more complex target attribute of economic value. Other examples of radical insensitivity to scope lend themselves to a similar interpretation. Among others, Kahneman (1986) found that Toronto residents were willing to pay almost as much to clean up polluted lakes in a small region of Ontario as to clean up all the polluted lakes in Ontario, and McFadden and Leonard (1993) reported that residents in four western states were willing to pay only 28% more to protect 57 wilderness areas than to protect a single area (for more discussion of scope insensitivity, see Frederick & Fischhoff, 1998).

The similarity between WTP statements and categorical predictions is not limited to such demonstrations of almost complete extension neglect. The two responses also yield similar results when extension and prototype information are varied factorially within subjects. Figure 12.2(a) shows the results of a study of WTP for programs that prevented different levels of damage to species of varying popularity (Ritov & Kahneman, unpublished observations, cited in Kahneman, Ritov, & Schkade, 1999). As in the case of base rate [Figure 12.2(c)], extensional information (levels of damage) combines additively with nonextensional information. This rule of combination is unreasonable; in any plausible theory of value, the lines would fan out.

Finally, the role of the emotion evoked by a prototypical instance was also examined directly in the same experiment, using the heuristic elicitation paradigm introduced earlier: Some respondents were asked to imagine that they saw a television program documenting the effect of adverse ecological circumstances on individual members of different species. The respondents indicated, for each species, how much concern they expected to feel while watching such a documentary. The correlation between this measure of affect and willingness to pay, computed across species, was .97.

Duration Neglect in the Evaluation of Experiences

We next discuss experimental studies of the global evaluation of experiences that extend over some time, such as a pleasant or a horrific film clip (Fredrickson & Kahneman, 1993), a prolonged unpleasant noise (Schreiber & Kahneman, 2000), pressure from a vise (Ariely, 1998), or a painful medical procedure (Redelmeier & Kahneman, 1996). Participants in these studies provided a continuous or intermittent report of hedonic or affective state, using a designated scale of momentary affect (Figure 12.3). When the episode had ended, they indicated a global evaluation of "the *total* pain or discomfort" associated with the entire episode.

We first examine the normative rules that apply to this task. The global evaluation of a temporally extended outcome is an extensional attribute, which is governed by a distinctive logic. The most obvious rule is temporal monotonicity: There is a compelling intuition that adding an extra period of pain to an episode of discomfort can only make it worse overall. Thus, there are two ways of making a bad episode worse – making the discomfort more intense or prolonging it. It must therefore be possible to trade off intensity against duration. Formal analyses have identified conditions under which the total utility of an episode is equal to the temporal integral of a suitably transformed measure of the instantaneous utility associated with each moment (Kahneman, 2000a; Kahneman, Wakker, & Sarin, 1997).

Next, we turn to the psychology. Fredrickson and Kahneman (1993) proposed a "snapshot model" for the retrospective evaluation of episodes, which again involves two acts of substitution: First, the episode is represented by a prototypical moment; next, the affective value attached to the representative moment is substituted for the extensional target attribute of global evaluation. The snapshot model was tested in an experiment in which participants provided continuous ratings of their affect while watching plotless films that varied in duration and affective value (e.g., fish swimming in coral reefs, pigs being beaten to death with clubs), and later reported global evaluations of their experiences. The central finding was that the retrospective evaluations of these observers were predicted with substantial accuracy by a simple average of the peak affect recorded during a film and the end affect reported as the film was about to end. This has been called the peak/end rule. However, the correlation between retrospective evaluations and the duration of the films was negligible – a finding that Fredrickson and Kahneman labeled *duration neglect*. The resemblance of duration neglect to the neglect of scope and base rate is striking and unlikely to be accidental. In this analysis, all three are manifestations of extension neglect caused by the use of a prototype heuristic.

The peak/end rule and duration neglect have both been confirmed on multiple occasions. Figure 12.3 presents raw data from a study reported by Redelmeier and Kahneman (1996), in which patients undergoing colonoscopy reported their current level of pain every 60 seconds throughout the procedure. Here again, an average of peak and end pain quite accurately predicted subsequent global evaluations and choices. The duration of the procedure varied considerably among patients (from 4 to 69 minutes), but these differences were not reflected in subsequent global evaluations in accord with duration neglect. The implications of these psychological rules of evaluation are paradoxical. In Figure 12.3, for example, it appears evident that patient B had a worse colonoscopy than patient A (on the assumption they used the scale similarly). However, it is also apparent that the peak/end average was worse for patient A, whose procedure ended at a moment of relatively intense pain. The peak/end rule prediction for these two profiles is that patient A would evaluate the procedure more negatively than patient B and would be more likely to prefer to undergo a barium enema rather than a repeat colonoscopy. The prediction was correct for these two individuals and confirmed by the data of a large group of patients.

The effects of substantial variations of duration remained small (although statistically robust) even in studies conducted in a factorial design. Figure 12.2(d) is drawn from a study of responses to ischemic pain (Ariely, 1998), in which duration varied by a factor of 4. The peak/end average accounted for 98%

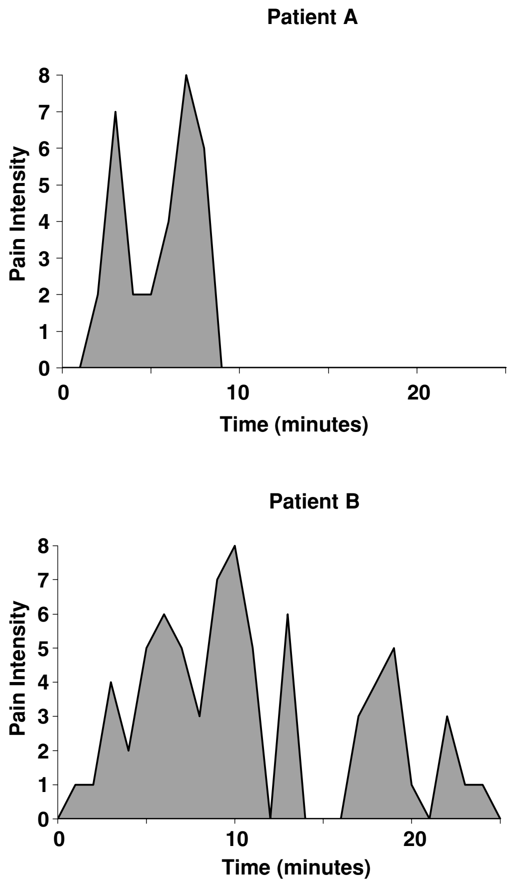

Figure 12.3. Pain intensity reported by two colonoscopy patients.

of the systematic variance of global evaluations in that study and for 88% of the variance in a similar factorial study of responses to loud unpleasant sounds [Schreiber & Kahneman, 2000, Figure 12.2(b)]. Contrary to the normative standard for an extensional attribute, the effects of duration and of other determinants of evaluation were additive [Figures 12.2(b) and 12.2(d)].

The participants in these studies were well aware of the relative duration of their experiences and did not consciously decide to ignore duration in their evaluations. As Fredrickson and Kahneman (1993) noted, duration neglect is an attentional phenomenon:

> *...duration neglect does not imply that duration information is lost, nor that people believe that duration is unimportant...people may be aware of duration and consider it important in the abstract [but] what comes most readily to mind in evaluating episodes are the salient moments of those episodes and the affect associated with those moments. Duration neglect might be overcome, we suppose, by drawing attention more explicitly to the attribute of time. (p. 54)*

This comment applies equally well to other instances of extension neglect: The neglect of base rate in categorical prediction, the neglect of scope in willingness to pay, the neglect of sample size in evaluations of evidence (Griffin & Tversky, 1992; Tversky & Kahneman, 1971), and the neglect of probability of success in evaluating a program of species preservation (DeKay & McClelland, 1995). More generally, inattention plays a similar role in any situation in which the intuitive judgments generated by system 1 violate rules that would be accepted as valid by the more deliberate reasoning that we associate with system 2. As we noted earlier, the responsibility for these judgmental mishaps is properly shared by the two systems: System 1 produces the initial error, and system 2 fails to correct it, although it could.

Violations of Dominance

The conjunction fallacy observed in the Linda problem is an example of a dominance violation in judgment: Linda must be at least as likely to be a bank teller as to be a feminist bank teller, but people believe the opposite. Insensitivity to extension (in this case, base rate) effectively guarantees the existence of such dominance violations. For another illustration, consider the question: "How many murders were there last year in [Detroit/Michigan]?" Although there cannot be more murders in Detroit than in Michigan, because Michigan contains Detroit, the word "Detroit" evokes a more violent image than the word "Michigan" (except of course for people who immediately think of Detroit when Michigan is mentioned). If people use an impression of violence as a heuristic and neglect geographic extension, their estimates of murders in the city may exceed their estimates for the state. In a large sample of University of Arizona students, this hypothesis was confirmed – the median estimate of the number of murders was 200 for Detroit and 100 for Michigan.

Violations of dominance akin to the conjunction fallacy have been observed in several other experiments involving both indirect (between-subjects) and direct tests. In a clinical experiment reported by Redelmeier, Katz, and Kahneman (2001), half of a large group of patients ($N = 682$) undergoing a colonoscopy were randomly assigned to a condition that made the actual experience strictly worse. Unbeknownst to the patient, the physician deliberately delayed the removal of the colonoscope for approximately 1 minute beyond the normal time. The instrument was not moved during the extra period. For many patients, the mild discomfort of the added period was an improvement relative to the pain than they had just experienced. For these patients, of course, prolonging the procedure reduced the peak/end average of discomfort. As expected, retrospective evaluations were less negative in the experimental group, and a 5-year follow-up showed that participants in that group were also somewhat more likely to comply with recommendations to undergo a repeat colonoscopy (Redelmeier, Katz, & Kahneman, 2001).

In an experiment that is directly analogous to the demonstrations of the conjunction fallacy, Kahneman et al. (1993) exposed participants to two cold-pressor experiences, one with each hand: a "short" episode (immersion of one hand in 14 °C water for 60 seconds), and a "long" episode (the short episode, plus an additional 30 seconds during which the water was gradually warmed to 15 °C). The participants indicated the intensity of their pain throughout the experience. When they were later asked which of the two experiences they preferred to repeat, a substantial majority chose the long trial. These choices violate dominance, because after 60 seconds in cold water anyone will prefer the immediate experience of a warm towel to 30 extra seconds of slowly diminishing pain. In a replication, Schreiber and Kahneman (2000, experiment 2) exposed participants to pairs of unpleasant noises in immediate succession. The participants listened to both sounds and chose one to be repeated at the end of the session. The "short" noise lasted 8 seconds at 77 db. The "long" noise consisted of the short noise plus an extra period (of up to 24 seconds) at 66 db (less aversive, but still unpleasant and certainly worse than silence). Here again, the longer noise was preferred most of the time, and this unlikely preference persisted over a series of five choices.

The violations of dominance in these direct tests are particularly surprising because the situation is completely transparent. The participants in the experiments could easily retrieve the durations of the two experiences between which they had to choose, but the results suggest that they simply ignored duration. A simple explanation is that the results reflect "choosing by liking" (see Frederick, 2002). The participants in the experiments simply followed the normal strategy of choice: "When choosing between two familiar options, consult your retrospective evaluations and choose the one that you like most (or dislike least)." Liking and disliking are products of system 1, which do not conform to the rules of extensional logic. System 2 could have intervened, but in these experiments it generally did not. Kahneman et al. (1993) described a participant in their study, who chose to repeat the long cold-pressor experience. Soon after the choice was recorded, the participant was asked which of the two experiences was longer. As he correctly identified the long trial, the participant was heard to mutter "the choice I made doesn't seem to make much sense." Choosing by liking is a form of mindlessness (Langer, 1978), which illustrates the casual governance of system 2.

Like the conjunction fallacy in direct tests, which we discussed earlier, violations of temporal monotonicity in choices should be viewed as an expendable flourish. Because the two aversive experiences occurred within a few minutes of each other and respondents could accurately recall the duration of the two events, system 2 had enough information to override choosing by liking. Its failure to do so is analogous to the failures observed in direct tests of the Linda problem. In both cases, the violations of dominance tell us nothing new about system 1; they only illustrate an unexpected weakness of system 2. Just as the theory of intuitive categorical prediction would have remained intact if the conjunction fallacy had not "worked" in a direct test, the model of evaluation by moments would have survived even if violations of dominance had been eliminated in highly transparent situations. The same methodological issues arise in both contexts. Between-subjects experiments or subtle tests are most appropriate for studying the basic intuitive evaluations of system 1, and also most likely to reveal complete extension neglect. Factorial designs in which extension is manipulated practically guarantee an effect of this variable, and almost guarantee that it will be additive, as in Figures 12.2(b) and 12.2(d) (Ariely, 1998; Ariely, Kahneman, & Loewenstein, 2000; Schreiber & Kahneman, 2000). Finally, although direct choices sometimes yield systematic violations of dominance, these violations can be avoided by manipulations that prompt system 2 to take control.

In our view, the similarity of the results obtained in diverse contexts is a compelling argument for a unified interpretation, and a significant challenge to critiques that pertain only to selected subsets of this body of evidence. A number of commentators have offered competing interpretations of base rate neglect (Cosmides & Tooby, 1996; Koehler, 1996), insensitivity to scope in WTP (Kopp, 1992), and duration neglect (Ariely & Loewenstein, 2000). However, these interpretations are generally specific to a particular task and would not carry over to analogous findings in other domains. Similarly, the various attempts to explain the conjunction fallacy as an artifact do not explain analogous violations of dominance in the cold-pressor experiment. The account we have offered is, in contrast, equally applicable to all three contexts and possibly others (see also Kahneman, Ritov, & Schkade, 1999). We attribute extension neglect and violations of dominance to a lazy system 2, and to a prototype heuristic that combines two processes of system 1: the representation of categories by prototypes and the substitution of a nonextensional heuristic attribute for an extensional target attribute. We also propose that people have some appreciation of the role of extension in the various judgment tasks. Consequently, they will incorporate extension in their judgments when their attention is drawn to this factor – most reliably in factorial experiments, and sometimes (although not always) in direct tests. The challenge for competing interpretations is to provide a unified account of the diverse phenomena that have been considered in this section.

Conclusions and Future Directions

The original goal of the heuristics and biases program was to understand intuitive judgment under uncertainty. Heuristics were described as a collection of disparate cognitive procedures, related only by their common function in a particular judgmental domain – choice under uncertainty. It now appears, however, that judgment heuristics are applied in a wide variety of domains and share a common process of *attribute substitution*, in which difficult judgments are made by substituting conceptually or semantically related assessments that are simpler and more readily accessible.

The current treatment explicitly addresses the conditions under which intuitive judgments are modified or overridden. Although attribute substitution provides an initial input into many judgments, it need not be the sole basis for them. Initial impressions are often supplemented, moderated, or overridden by other considerations, including the recognition of relevant logical rules and the deliberate execution of learned algorithms. The role of these supplemental or alternative inputs depends on characteristics of the judge and the judgment task.

Our use of the dual-process terminology does not entail a belief that every mental operation (including each postulated heuristic) can be definitively assigned to one system or the other. The placement of dividing lines between "systems" is arbitrary because the bases by which we characterize mental operations (difficulty of acquisition, accessibility to introspection, and disruptability) are all continua. However, this does not make distinctions less meaningful; there is broad agreement that mental operations range from rapid, automatic, perception-like impressions to deliberate computations that apply explicit rules or external aids.

Many have questioned the usefulness of the notion of heuristics and biases by pointing to inconsistencies in the degree to which illusions are manifested across different studies. However, there is no mystery here to explain. Experimental studies of "the same" cognitive illusions can yield different results for two reasons: (1) because of variation in factors that determine the accessibility of the intuitive illusion, and (2) because they vary in factors that determine the accessibility of the corrective thoughts that are associated with system 2. Both types of variation can often be anticipated because of the vast amount of psychological knowledge that has accumulated about the different sets of factors that determine the ease with which thoughts come to mind – from principles of grouping in perception to principles that govern transfer of training in rule learning (Kahneman, 2003). Experimental surprises will occur, of course, and should lead to refinements in the understanding of the rules of accessibility.

The argument that system 1 will be expressed unless it is overridden by system 2 sounds circular, but it is not, because empirical criteria can be used to test whether a particular characterization of the two systems is accurate. For example, a feature of the situation will be associated with system 2 if it is shown to influence judgments only when attention is explicitly directed to it (through, say, a within-subjects design). In contrast, a variable will be associated with system 1 if it can be shown to influence even those judgments that are made in a split second. Thus, one need not be committed, a priori, to assigning a process to a particular system; the data will dictate the best characterization.

The two-system model is a framework that combines a set of empirical generalizations about cognitive operations with a set of tests for diagnosing the types of cognitive operations that underlie judgments in specific situations. The generalizations and the specific predictions are testable and can be recognized as true or false. The framework itself will be judged by its usefulness as a heuristic for research.

Acknowledgments

This chapter is a modified version of a chapter by Kahneman and Frederick (2002). Preparation of this chapter was supported by grant SES-0213481 from the National Science Foundation.

Note

1. The entries plotted in Figure 12.1 are averages of multiple judgments, and the correlations are computed over a set of judgment objects. It should be noted that correlations between averages are generally much higher than corresponding correlations within the data of individual respondents (Nickerson, 1995). Indeed, group results may even be unrepresentative if they are dominated by a few individuals who produce more variance than others and have an atypical pattern of responses. Fortunately, this particular hypothesis is not applicable to the experiments of Figure 12.1, in which all responses were ranks.

References

Agnoli, F. (1991). Development of judgmental heuristics and logical reasoning: Training counteracts the representativeness heuristic. *Cognitive Development*, 6, 195–217.

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology*, *21*, 515–550.

Anderson, N. H. (1991). *Contributions to information integration theory. Vol. I: cognition*. Hillsdale, NJ: Erlbaum.

Anderson, N. H. (1996). *A functional theory of cognition*. Mahwah, NJ: Erlbaum.

Ariely, D. (1998). Combining experiences over time: The effects of duration, intensity changes and on-line measurements on retrospective pain evaluations. *Journal of Behavioral Decision Making*, *11*, 19–45.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Ariely, D., Kahneman, D., & Loewenstein, G. (2000). Joint comment on "When does duration matter in judgment and decision making?" (Ariely & Loewenstein, 2000). *Journal of Experimental Psychology: General*, *129*, 524–529.

Ariely, D., & Loewenstein, G. (2000). When does duration matter in judgment and decision making. *Journal of Experimental Psychology: General*, *129*, 508–523.

Ayton, P. (1998). How bad is human judgment? In G. Wright & P. Goodwin (Eds.), *Forecasting with judgment*. West Sussex, UK: Wiley.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *Advances in social cognition* (Vol. 10). Mahwah, NJ: Erlbaum.

Bargh, J. A., Bond, R. N., Lombardi, W. J., & Tota, M. E. (1986). The additive nature of chronic and temporary sources of construct accessibility. *Journal of Personality and Social Psychology*, 50(5), 869–878.

Bar-Hillel, M., & Fischhoff, B. (1981). When do base-rates affect predictions? *Journal of Personality and Social Psychology*, *41* (4), 671–680.

Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. *Journal of Personality and Social Psychology*, *41* (4), 671–680.

Begg, I., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121* (4), 446–458.

Begg, I., & Armour, V. (1991). Repetition and the ring of truth: Biasing comments. *Canadian Journal of Behavioural Science*, 23, 195–213.

Birnbaum, M. H. (1999). How to show that 9 > 221: Collect judgments in a between-subjects design. *Psychological Methods*, *4*(3), 243–249.

Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, 45, 792–804.

Bless, H., Clore, G. L., Schwarz, N., Golisano, V., Rabe, C., & Wolk, M. (1996). Mood and the use of scripts: Does a happy mood really lead to mindlessness? *Journal of Personality and Social Psychology*, *71* (4), 665–679.

Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science*, *1* (5), 319–322.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review*, 50, 255–272.

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology*, 65, 5–17.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.

Chapman, G., & Johnson, E. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. Griffin, & D. Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 120–138). New York: Cambridge University Press.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. San Diego: Academic Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Cosmides L., & Tooby J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58(1), 1–73.

DeKay, M. L., & McClelland, G. H. (1995). Probability and utility components of endangered species preservation programs. *Journal of Experimental Psychology: Applied*, *2*, 60–83.

Desvouges, W. H., Johnson, F., Dunford, R., Hudson, S., Wilson, K., & Boyle, K. (1993). Measuring resource damages with contingent valuation: Tests of validity and reliability.

In *Contingent valuation: A critical assessment*. Amsterdam: North Holland.

Diamond, P. (1996). Testing the internal consistency of contingent valuation surveys. *Journal of Environmental Economics and Management*, 28, 155–173.

Dulany, D. E. & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, 9, 85–110.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49(8), 709–724.

Evans, J., Handley, S. J., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*, 77, 197–213.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, 13, 1–17.

Frederick, S. W. (2002). Automated choice heuristics. In T. Gilovich, D. Griffin, & D. Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 548–558). New York: Cambridge University Press.

Frederick, S. W. (2004). *(Im)patient, (un)adventurous, (in)sensitive, and (im)moral: Cognitive (in)ability as a predictor of time preference, risk preference, magnitude sensitivity, and cheating*. Working paper, Massachusetts Institute of Technology, Cambridge, MA.

Frederick, S. W., & Fischhoff, B. (1998). Scope (in)sensitivity in elicited valuations. *Risk, Decision, and Policy*, 3, 109–123.

Frederick, S. W., & Nelson, L. (2004). *Attribute substitution in physical judgments*, Working paper, Massachusetts Institute of Technology, Cambridge, MA.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, 65(1), 45–55.

Ganzach, Y., & Krantz, D. H. (1990). The psychology of moderate prediction: I. Experience with multiple determination. *Organizational Behavior and Human Decision Processes*, 47, 177–204.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 83–115). Chichester, UK: Wiley.

Gigerenzer, G., Czerlinski, J., & Martignon, L. (2002). How good are fast and frugal heuristics? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 559–581). New York: Cambridge University Press.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103(4), 650–669.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content – The use of base rates as a continuous variable. *Journal of Experimental Psychology – Human Perception and Performance*, 14(3), 513–525.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684–704.

Gigerenzer, G., Todd, P. M., & the ABC Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gilbert, D. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.

Gilbert, D. (1991). How mental systems believe. *American Psychologist*, 46(2), 107–119.

Gilbert, D. (1999). What the mind's not. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 3–11). New York: Guilford Press.

Gould, S. J. (1991). *Bully for brontosaurus. Reflections in natural history*. New York: Norton.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, 24(3), 411–435.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, 62, 255–262.

Hammond, K. R. (1996). *Human judgment and social policy*. New York: Oxford University Press.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review*, 51, 358–374.

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, 12(4), 275–305.

Higgins, E. T., & Brendl, C. M. (1995). Accessibility and applicability: Some "activation rules" influencing judgment. *Journal of Experimental Social Psychology*, 31, 218–243.

Hilton, D. J., & Slugoski, B. R. (2001). Conversational processes in reasoning and explanation. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology. Vol. 1: Intraindividual processes* (pp. 181–206). Oxford, UK: Blackwell.

Isen, A. M., Nygren, T. E., & Ashby, F. G. (1988). Influence of positive affect on the subjective utility of gains and losses – It is just not worth the risk. *Journal of Personality and Social Psychology*, 55(5), 710–717.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 3, 306–340.

Kahneman, D. (1986). Valuing environmental goods: An assessment of the contingent valuation method. In R. Cummings, D. Brookshire, & W. Schulze (Eds.), *Valuing environmental goods: An assessment of the contingent valuation method*. Totowa, NJ: Rowman and Allanheld.

Kahneman, D. (2000a). Experienced utility and objective happiness: A moment-based approach. In D. Kahneman & A. Tversky (Eds.), *Choices, values, and frames*. New York: Cambridge University Press and the Russell Sage Foundation.

Kahneman, D. (2000b). A psychological point of view: Violations of rational rules as a diagnostic of mental processes (commentary on Stanovich and West). *Behavioral and Brain Sciences*, 23, 681–683.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697–720.

Kahneman, D., & Frederick, (2002). In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics of intuitive judgment*. New York: Cambridge University Press.

Kahneman, D., Fredrickson, D. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, 4, 401–405.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality with its alternatives. *Psychological Review*, 93, 136–153.

Kahneman, D., & Ritov, I. (1994). Determinants of stated willingness to pay for public goods: A study in the headline method. *Journal of Risk and Uncertainty*, 9, 5–38.

Kahneman, D., Ritov, I., & Schkade, D. (1999). Economic preferences or attitude expressions? An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty*, 19, 203–235.

Kahneman, D., Schkade, D. A., & Sunstein, C. R. (1998). Shared outrage and erratic awards: The psychology of punitive damages. *Journal of Risk and Uncertainty*, 16, 49–86.

Kahneman, D., Slovic, P., & Tversky, A. E. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, 11, 123–141.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, 103, 582–591.

Kahneman, D., & Varey, C. A. (1990). Propensities and counterfactuals: The loser that almost won. *Journal of Personality and Social Psychology*, 59(6), 1101–1110.

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics*, 112, 375–405.

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language*, 28, 375–385.

Koehler, J. (1996). The base-rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, 19, 1–53.

Kopp, R. (1992). Why existence value should be used in cost–benefit analysis. *Journal of Policy Analysis and Management*, 11, 123–130.

Krosnick, J. A., Li, F., & Lehman, D. R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgment. *Journal of Personality and Social Psychology*, 59, 1140–1152.

Langer, E. J. (1978). Rethinking the role of thought in social interaction. In Harvey, Ickes,

& Kidd (Eds.), *New directions in attribution research* (Vol. 2). Potomac, MD: Erlbaum.

Macchi, L. (1995). Pragmatic aspects of the base-rate fallacy. *Quarterly Journal of Experimental Psychology*, *48A*, 188–207.

Mandler, G., Hamson, C., & Dorfman, J. (1990). Tests of dual process theory – Word priming and recognition. *Quarterly Journal of Experimental Psychology*, *42*(4), 713–739.

Margolis, H. (1987). *Patterns, thinking & cognition*. Chicago: University of Chicago Press.

McFadden, D., & Leonard, G. K. (1993). Issues in the contingent valuation of environmental goods: Methodologies for data collection and analysis. In Hausman (Ed.), *Contingent valuation. A critical assessment*. Amsterdam: North-Holland.

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, *12*, 269–275.

Michotte, A. (1963). *The perception of causality*. New York: Basic Books.

Mitchell, R., & Carson, R. (1989). *Using surveys to value public goods: The contingent valuation method*. Washington, DC: Resources for the Future.

Nickerson, C. (1995). Does willingness-to-pay reflect the purchase of moral satisfaction? A reconsideration of Kahneman and Knetsch. *Journal of Environmental Economics and Management*, *28*, 126–133.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*(4), 339–363.

Novemsky, N., & Kronzon, S. (1999). How are base-rates used, when they are used: A comparison of Bayesian and additive models of base-rate use. *Journal of Behavioral Decision Making*, *12*, 55–69.

Ordóñez, L., & Benson, L., III. (1997). Decisions under time pressure: How time constraint affects risky decision making. *Organizational Behavior and Human Decision Processes*, *71*(2), 121–140.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*, 407–418.

Pashler, H. E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Poulton, (1989). *Bias in quantifying judgments*. London: Erlbaum.

Redelmeier, D., Katz, J., & Kahneman, D. (2001). Memories of colonoscopy: A randomized trial.

Redelmeier, D., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*, 3–8.

Schreiber, C., & Kahneman, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology: General*, *129*, 27–42.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Erlbaum.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*, 195–202.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*(3), 513–523.

Schwarz, N., Strack, F., Hilton, D., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition*, *9*, 67–84.

Schwarz, N., & Vaughn, L. A. (2002). The availability heuristic revisited: Ease of recall and content of recall as distinct sources of information. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics & biases: The psychology of intuitive judgment* (pp. 103–119). New York: Cambridge University Press.

Shweder, R. (1977). Likeness and likelihood in everyday thought – Magical thinking in judgments about personality. *Current Anthropology*, *18*(4), 637–658.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, *61*, 394–403.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics & biases: The psychology of intuitive judgment* (pp. 379–396). New York: Cambridge University Press.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics & biases: The psychology of intuitive judgment* (pp. 397–420). New York: Cambridge University Press.

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Stanovich, K. E., & West, R. (2002). Individual differences in reasoning: Implications for the rationality debate? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics & biases: The psychology of intuitive judgment* (pp. 421–440). New York: Cambridge University Press.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64, 153–181.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.

Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communication: The social determinants of information use in judgments of life-satisfaction. *European Journal of Social Psychology*, 18, 429–442.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76, 105–110.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–98). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, 59, S251–S278.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129–140.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151–175.

Zajonc, R. B. (1997). Emotions. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., pp. 591–632). New York: Oxford University Press.